



COMPUTATIONAL  
AND STRUCTURAL  
BIOTECHNOLOGY  
JOURNAL

journal homepage: [www.elsevier.com/locate/csbj](http://www.elsevier.com/locate/csbj)



## Mini Review

# Mechanisms and impact of genetic recombination in the evolution of *Streptococcus pneumoniae*

Chrispin Chaguza<sup>\*</sup>, Jennifer E. Cornick<sup>1</sup>, Dean B. Everett<sup>1</sup>

Department of Clinical Infection, Microbiology and Immunology, Institute of Infection and Global Health, University of Liverpool, L69 7BE Liverpool, UK  
Microbes, Immunity and Vaccines Theme, Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Queen Elizabeth Central Hospital, P/Bag 30096, Blantyre, Malawi

## ARTICLE INFO

### Article history:

Received 2 February 2015  
Received in revised form 27 March 2015  
Accepted 31 March 2015  
Available online 8 April 2015

### Keywords:

Genetic recombination  
*Streptococcus pneumoniae*  
Evolution

## ABSTRACT

*Streptococcus pneumoniae* (the pneumococcus) is a highly recombinogenic bacterium responsible for a high burden of human disease globally. Genetic recombination, a process in which exogenous DNA is acquired and incorporated into its genome, is a key evolutionary mechanism employed by the pneumococcus to rapidly adapt to selective pressures. The rate at which the pneumococcus acquires genetic variation through recombination is much higher than the rate at which the organism acquires variation through spontaneous mutations. This higher rate of variation allows the pneumococcus to circumvent the host innate and adaptive immune responses, escape clinical interventions, including antibiotic therapy and vaccine introduction. The rapid influx of whole genome sequence (WGS) data and the advent of novel analysis methods and powerful computational tools for population genetics and evolution studies has transformed our understanding of how genetic recombination drives pneumococcal adaptation and evolution. Here we discuss how genetic recombination has impacted upon the evolution of the pneumococcus.

© 2015 Chaguza et al. Published by Elsevier B.V. on behalf of the Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Contents

1. Introduction . . . . .	241
2. Genetic Exchange in Pneumococcal Genomes . . . . .	242
3. Core, Accessory and Pan-genome of <i>S. pneumoniae</i> . . . . .	243
4. Capturing Genetic Recombination Signals . . . . .	243
5. Contributions of Recombination and Random Mutations . . . . .	243
6. Pneumococcal Virulence and Antibiotic Resistance . . . . .	244
7. Recombination Drives Vaccine Escape . . . . .	245
8. Conclusions . . . . .	245
Acknowledgements . . . . .	246
References . . . . .	246

## 1. Introduction

Once described as the ‘Captain of the men of death’ by Sir William Osler [1], *Streptococcus pneumoniae* (the pneumococcus) is a gram-

positive, aerobic, human-adapted commensal that colonises the human nasopharynx [2]. Pneumococcal carriage rates range from 10 to 80% [3,4]. The wide range of carriage rates is mainly associated with age, with children exhibiting higher carriage than adults [3]. Pneumococcal carriage precedes spread to sterile parts of the body, which results in non-invasive diseases such as otitis media, and invasive pneumococcal diseases (IPD); bacteraemia, pneumonia and meningitis [5,6]. Pneumococcal infections are responsible for approximately one million deaths in children under five annually, of which more than 90% are in resource poor settings, particularly in Sub-Saharan Africa, Latin America and Asia

<sup>\*</sup> Corresponding author at: Department of Clinical Infection, Microbiology and Immunology, Institute of Infection and Global Health, University of Liverpool, L69 7BE Liverpool, UK.

E-mail address: [Chrispin.Chaguza@liverpool.ac.uk](mailto:Chrispin.Chaguza@liverpool.ac.uk) (C. Chaguza).

<sup>1</sup> These authors made equal contribution.

[7]. In addition to geography, additional factors including, age, smoking and co-infection with other diseases such as HIV predispose individuals to pneumococcal infections [8]. At least 93 pneumococcal serotypes are known, based on the structure and antigenicity of the pneumococcal polysaccharide capsule [9–12], a major virulence factor that surrounds the pneumococcal cell wall [13]. Heterogeneity between pneumococcal serotypes has been reported both in terms of the invasive disease potential [14] and global geographical distribution [15]. Pneumococcal Conjugate Vaccines (PCVs) that target the pneumococcal capsule offer protection against the serotypes most commonly associated with invasive disease. PCV7 consists of serotypes 4, 6B, 9V, 14, 18C and 23F whereas the PCV10 consists of the same set of serotypes in PCV7 with the addition of serotypes 1, 5 and 19F. Both PCV7 and PCV10 have been shown to have high efficacy against the vaccine type (VT) serotypes they incorporate and initially led to a dramatic decline in IPD [16,17]. However, in the aftermath of the PCV7 rollout, several studies have reported an increase in the rates of non-vaccine type (NVT) serotypes in carriage [18] and invasive disease [19–22]. The newly introduced PCV13 vaccine, which contains all the serotypes in PCV10 plus 19A, 6B, 7F, promises to further reduce the overall IPD burden by targeting additional serotypes some of which increased in prevalence after the rollout of the previous formulations [17,23].

For *S. pneumoniae* to survive, it is essential that it is able to rapidly adapt to clinical interventions and its human hosts' immune responses [24]. The introduction of genomic variation is one of the main mechanisms it can employ to adapt to the host environment, avoid the host immune response and evade clinical interventions. Genetic variation arises in pneumococci and other bacterial genomes as a result of errors by DNA polymerase and the DNA mismatch repair (MMR) system during DNA replication [25]. These alterations in the encoded nucleotide base sequence are known as random mutations. The rate at which random mutations occur varies between bacteria due to the differences in fidelity of DNA polymerases. Spontaneous mutations are either maintained or discarded in the population depending on the fitness cost of the new genotypes [26]. Advantageous genotypes with a lower fitness cost are likely to be favoured by positive Darwinian selection and can rapidly spread and become prevalent within a bacterial population. However, certain genotypes can increase in prevalence purely by linkage to other loci that are undergoing selective sweeps, a process known as genetic hitchhiking [27]. Furthermore within small populations certain genotypes are more likely to become more or less prevalent by chance alone regardless of the fitness costs, this is known as genetic drift. Genomic variation can also arise through lateral transfer of DNA fragments between bacteria followed by integration into the host cell genome, in a process known as genetic recombination or 'prokaryotic sex' [28,29]. Numerous studies have demonstrated the contributions of random mutation and recombination to the evolution of bacteria [24,30–33]. However, variations exist in their relative importance depending on the bacterial species or individual strain considered [34]. Together, all the aforementioned processes play a role in the evolution and adaptation of the pneumococcus [35]. In recent years, the rapid generation of WGS data, coupled with the development of efficient analysis tools computational methods have helped to provide better insights into the patterns and dynamics of bacterial recombination.

In bacteria, uptake of exogenous DNA is achieved via three main processes; transformation, conjugation and transduction. Bacterial conjugation occurs when there is direct cell-to-cell contact of the bacteria exchanging DNA via a sex-pilus that protrudes from one cell into the other. During transduction, DNA is transferred from one bacterium to another by viruses that infect bacteria called bacteriophages. Bacterial transformation involves the acquisition of exogenous DNA from the bacterial surroundings followed by the integration of the acquired DNA into the host cell genome [36]. Depending on the diversity between donor and recipient DNA, this can result in the creation of variable or mosaic regions within the chromosomes that exhibit

incongruent evolutionary histories with other loci within the chromosomes that originate from multiple strains. In the pneumococcus, uptake of exogenous DNA predominantly occurs through transformation followed by integration into the host cell genome (recombination) [36,37]. DNA uptake is triggered by competence stimulating peptides (CSP) or exported pheromones that work in a quorum-sensing manner [38]. Quorum sensing involves regulation of gene expression in response to changes in the cell-population density [39]. Induction of competence after DNA damage is considered as an alternative to the SOS system, which is present in other species such as *Bacillus subtilis*, but absent in the pneumococcus [40]. Development of competence and transformation occurs during logarithmic growth of *S. pneumoniae* and requires the expression of key competence (com) genes [41,42]. These include *comC* gene that encodes a CSP and the two-component system (TCS) consisting of a histidine kinase (HK) (*comD*) and a cytoplasmic cognate response regulator (RR) (*comE*) [43]. The TCS activates *comX*, which activates the transcription of a cascade of late competence genes in the pneumococcus [44]. Other dispensable competence genes also help competent pneumococcal cells to outcompete the non-competent cells though triggering autolytic enzymes that kill them [45]. Competent pneumococci can thus acquire DNA released from the killed cells, a process known as fratricide [45]. Conjugative transfers of plasmids have also been documented, however, currently there is no evidence that suggest that phage-mediated transduction occurs in the pneumococcus [46].

Genetic recombination has been extensively studied in *S. pneumoniae* as a model organism. In this review, we discuss some of the recent findings on genetic recombination in the pneumococcus from published literature. We provide an overview of the main mechanisms of DNA uptake and exchange that facilitate pneumococcal recombination with other bacteria using WGS. We also discuss the potential biological and clinical impacts of recombination events in pneumococcal genomes with specific reference to antibiotic resistance, virulence and vaccine evasion.

## 2. Genetic Exchange in Pneumococcal Genomes

*S. pneumoniae* is a naturally competent bacterium, able to actively transport DNA fragments from the environment through the cell wall, into the cell cytoplasm. Transported DNA fragments can then be incorporated into the pneumococcal genome, a process known as genetic recombination [47]. In addition to pneumococcal competence, the simultaneous presence of both a donor and recipient bacterial strain is essential for genetic exchange to occur. Pneumococcal recombination has primarily been reported to occur during nasopharyngeal carriage, chronic polyclonal infection and biofilm formation [48,49]. The nasopharynx is a major reservoir for pneumococcal transmission and thus plays a role in disseminating recombinant bacterium across the human population. There are two main hypotheses to suggest why pneumococcal recombination has predominantly been observed in the nasopharynx. Firstly, the nasopharynx is an environment enriched with other microbial species, in addition to *S. pneumoniae*, which presents the opportunity for genetic exchanges. Secondly, the differential expression of the capsule has been suggested to play a significant role. A thinner capsule is expressed during nasopharyngeal carriage to facilitate attachment to the epithelial surface and this may indirectly allow for easy uptake of DNA as the pneumococcal capsule has been reported to inhibit recombination both *in vivo* and *in vitro* [50]. Additionally, comparative genomic analysis of over 3000 well sampled carriage pneumococci from a refugee camp in Thailand showed that unencapsulated pneumococci had a higher propensity to undergo genetic recombination than encapsulated pneumococci [51].

Lateral (or horizontal) gene transfer of exogenous DNA can result in recombination of either related or unrelated DNA segments. When transformation involves DNA exchange from closely related loci, it is known as homologous (legitimate) recombination. Homologous

recombination occurs in the core genome, a subset of genes that are shared and conserved across all members of the species under consideration. Homologous recombination also occurs between mobile genetic elements (MGE) such as insertion sequences (IS), integrons, bacteriophages, plasmids and transposons, considered being part of the accessory genome (non-core genome) [52]. Such recombination exchanges can occur between pneumococci or other closely related oral Streptococci including *Streptococcus mitis* and *Streptococcus oralis* [53]. Homologous recombination can also introduce new genes as exemplified by in molecular biology laboratories where synthetic gene constructs are inserted into the pneumococcal chromosomes [54]. When transformation occurs between unrelated loci, it is known as non-homologous (illegitimate) recombination. The term 'genetic recombination' will be used here to refer to all these forms of genetic exchanges.

### 3. Core, Accessory and Pan-genome of *S. pneumoniae*

Pneumococci possess a 2.1 megabases (Mb) pair circular genome that consists of over 2000 predicted protein coding regions and approximately 5% insertion elements [56]. The pneumococcus exhibits a high degree of genomic plasticity as evidenced by the level of genomic variability between isolates, with strains sharing approximately 74% identity at the nucleotide level [55]. On average the core genome of *S. pneumoniae* consists of 1647 predicted coding sequences (including paralogs) [55]. The remaining coding sequences that are not conserved in all members of the species are collectively referred as the 'accessory' genome, which usually contain dispensable genes that encode proteins that are not essential to the species. The total gene repertoire available to a species, the combined core and accessory genome, is termed the pan-genome [57]. *S. pneumoniae* consists of an open pan-genome which means that sequencing of new pneumococcal isolates continuously adds novel genes to the current gene pool. Variation in the pneumococcal core genome is predominantly introduced by random mutations and homologous recombination that involves both short and long stretches of DNA, whilst recombination involving unrelated loci is more restricted to the accessory genome. The accessory genome does not contain genes essential to cell survival however it plays an important role in bacterial pathogen evolution [58]. This is largely due to the acquisition of mobile genetic elements that harbour antibiotic resistance determinants and virulence factors [58].

### 4. Capturing Genetic Recombination Signals

Multiple approaches have been employed to identify occurrences of genetic recombination in bacterial genomes, reviewed elsewhere by Posada et al. [59]. These range from laboratory methods such as DNA hybridisation, to computational based methods such as Bayesian methods [24,55,60–62]. The Genealogies Unbiased by Recombination in Nucleotide Sequences (Gubbins) software was developed to identify recombination events in closely related pneumococci [63] but it has since been employed to study other bacterial species [64]. Closely related pneumococcal isolates belong to the same sequence clusters (SCs) or lineages. These SCs usually contains a single pneumococcal serogroup or clone inferred by multi-locus sequence typing (MLST). Within such similar isolates, the probability that a single nucleotide polymorphism (SNP) would occur at a specific genomic location, the 'background SNP density', is estimated as the total number of SNPs identified in the WGS divided by the overall size of the genome. Whole genome scans are used to determine genomic regions with statistically higher number of SNPs than would be expected by chance. This employs a sliding window approach that involves evaluation of a specified number of nucleotides across the genome. The SNP density within each sliding window is compared to the average background SNP density for the whole genome to determine regions with significantly higher number of SNPs than expected by chance alone [24,63]. Such atypical regions are most likely to have been acquired through

genetic recombination rather than spontaneous mutations, which on average introduces 2–4 novel mutations per genome per year. The amount of sequence diversity between donor and recipient strains determines the likelihood that genetic recombination will occur [65]. The true levels of recombination in *S. pneumoniae* and other pathogens are likely to be underestimated because some events are undetectable [59]. Such events occur between highly similar or identical loci and between distant species [59]. Ancient recombination events that involved distantly related taxa before their divergence into different species are difficult to detect by recombination algorithms because the signals in such loci maybe obscured due to the accumulation of additional point mutations [59].

### 5. Contributions of Recombination and Random Mutations

Beneficial mutations can sometimes arise independently in different bacterial strains. However, competition between strains and Darwinian selection can cause some beneficial mutations to be eliminated or become less prevalent in the bacterial population. This process is known as clonal interference. Genetic recombination provides a mechanism for sustaining such independent beneficial mutations from different strains through combining the different loci that contains the mutations, thus giving rise to recombinant strains that possess both mutations [66]. Although this process has been studied in *Escherichia coli* and *Saccharomyces cerevisiae*, it is presumptive that such mechanisms also occur in *S. pneumoniae* due to its highly recombinant nature [66,67]. The relative contributions of genetic recombination and random mutations to genomic diversifications of bacterial species have been previously reported [33]. A study that used the sequences of the seven MLST housekeeping genes to compare alleles introduced through random mutations and recombination events showed that the recombination to mutation (r/m) ratios varies between species. The r/m ratio measures the total number of SNPs imported from exogenous DNA through recombination (r) to those introduced randomly (m). For the pneumococcus, it was shown that a single nucleotide site is 50-fold more likely to change due to recombination than spontaneous point mutation [33]. Further studies using WGS have shown that genetic recombination in pneumococci is widespread, however, the r/m ratios reported were much lower (~7) than observed using MLST sequences [24]. Overall, these results suggest that nearly 90% of all polymorphisms in the pneumococcus have been introduced through recombination exchanges [24]. Further studies have also shown no significant variations in the mutation rates between pneumococcal SCs (lineages); an SC consists of a group of isolates with the same genetic backbone, which might not necessarily represent the same serotype [51]. Overall, an average of 2–4 mutations are introduced within a pneumococcal isolate per year regardless of the SC analysed in the phylogenies [51]. In contrast, every recombination event gives rise to an average of 72 mutations per isolate [24,32] but significant differences exist in levels of genetic recombination (r/m rates) observed between pneumococcal SCs [51].

The pneumococcal polysaccharide capsule is a major pneumococcal virulence factor. However as stated earlier, it is hypothesised to inhibit genetic recombination. New evidence from WGS analyses has shown that the rate of genetic recombination is higher among non-typeable (NT) pneumococci, which do not express a capsule, to encapsulated (typeable) pneumococci [51]. Analysis of pneumococcal strains within a single monophyletic clade consisting of serotype 14 isolates and NTs with the same genetic backbone as serotype 14, showed that the highest rates of genetic exchanges occurred in the NTs [51]. Pneumococcal isolates can undergo capsule switching whereby the serotype of a clone changes due to alteration in the capsule biosynthesis locus or through genetic recombination [68]. Capsule switching between encapsulated and unencapsulated states has been suggested as a transient state via which pneumococci can acquire rapid genetic diversity through recombination without the inhibitive effect of the surface capsule. The genetic

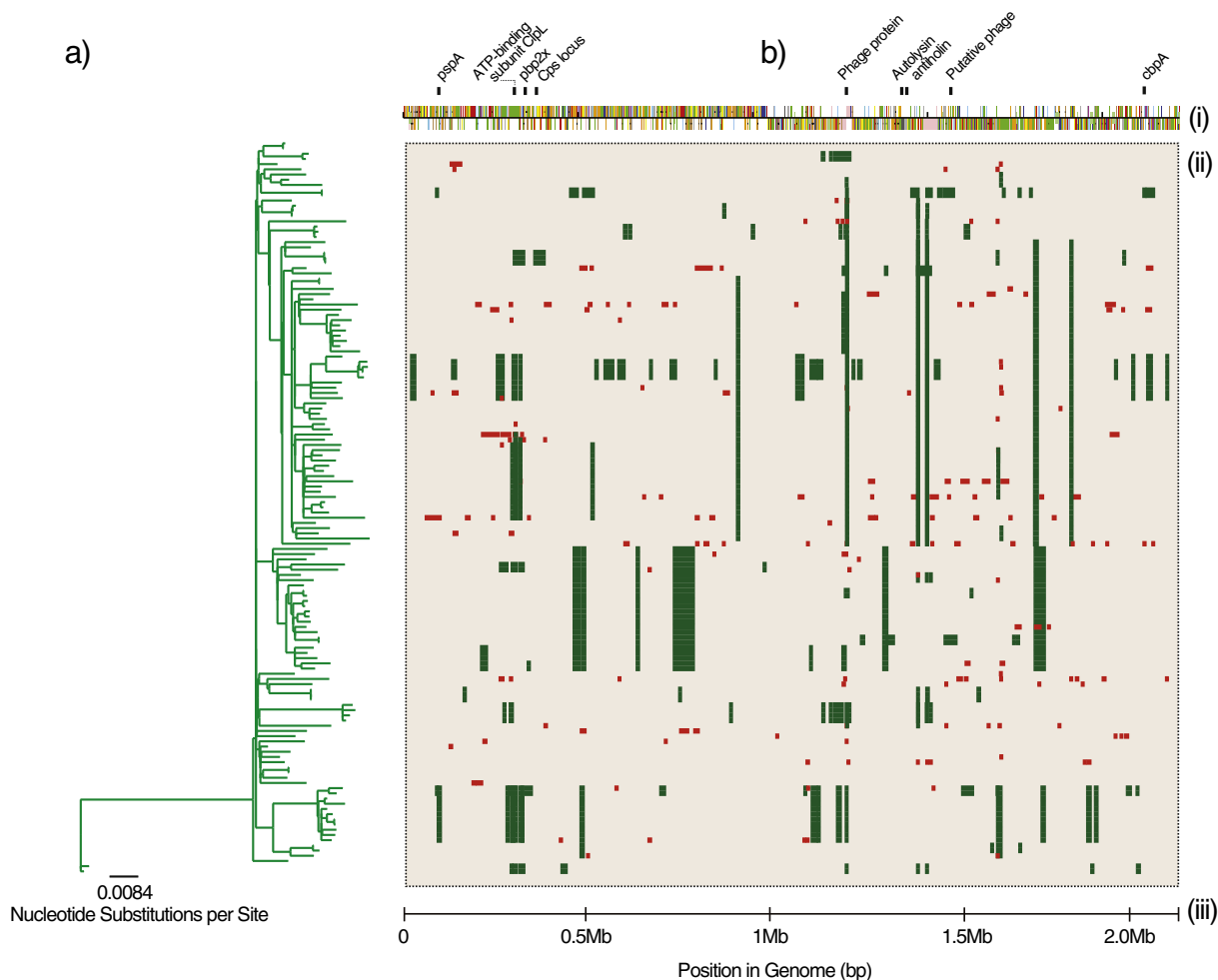
diversity acquired through this unencapsulated state would allow pneumococci to evade host immune responses or acquire novel resistance mechanisms that can lead to non-susceptibility to antibiotics upon return to the original or other capsular types [51,69]. Natural capsule switches mediated by recombination at the capsule polysaccharide synthesis (CPS) locus have previously been documented in pneumococci and preceded the introduction of the earliest pneumococcal vaccines [68,70].

Recombination events vary in size. Regardless of the pneumococcal serotype considered, genetic recombination events identified in the pneumococcus range from very small fragments to thousands of base pairs (bp) [68,71,72]. Two classes of recombination have been proposed based on the sizes of the recombination events: 1) micro-recombination which involve single and short stretches of DNA occur more frequently and 2) macro-recombination, which are usually rare and consists of multi-fragment replacements of DNA [72]. *In vitro* transformation of the pneumococcus followed by sequencing has shown that average recombination events are ~2 kilobases (Kb) [37]. Such *in vitro* experiments coupled with sequence analysis provide invaluable insight into the nature of recombination in the pneumococcus. Analysis of diverse collections of isolates at the population level has shown that on average,

the sizes of *in vivo* recombination events are higher than observed *in vitro* using the transformed mutants [24]. In a study of a single lineage of ST81 pneumococci (designated PMEN1 by the Pneumococcal Epidemiology Network) [73], recombination events were found to range from 3 bp to 72Kb with a mean of ~6Kb [24]. However, the mean size of recombinant blocks was higher in CC180 (serotype 3) lineage compared to PMEN1 indicating that the distribution of events varies by the lineage considered (Fig. 1) [71]. Overall, the sizes of recombination events in WGS follow a geometric distribution (specifically the exponential distribution) whereby short events are more prevalent than large recombination replacements. The large and multi-fragment recombination events (>30Kb) have been associated with major phenotype alterations such as capsule switching, which can also result in co-transfer of  $\beta$ -lactam resistance genes located near the CPS locus [68,72].

## 6. Pneumococcal Virulence and Antibiotic Resistance

A landmark experiment on pneumococcal transformation (recombination) by Fred Griffiths is the first and arguably most well-known study to demonstrate the biological importance of this process on



**Fig. 1.** Schematic representation of genetic recombination events in a subset of PMEN1 ST81 *S. pneumoniae* genomes from Croucher et al, (2011), Science. (a) A maximum likelihood phylogenetic tree generated using RAxML version 7.0.3 after removing recombination events in the multiple sequence alignment. (b) Location of genetic recombination events in *S. pneumoniae* genomes. (i) Schematic representation of the sequence features in the *S. pneumoniae* reference sequence. The genes in some regions with high density of recombination events are shown above the features. (ii) Rectangular matrix showing genomic regions where hypothetical genetic recombination events occurred. Recombination events were inferred as regions with atypically high density of polymorphisms. Each horizontal track beginning from the tips of the phylogeny represents a single aligned whole genome pneumococcal sequence. Green blocks shows the presence of a recombination event that was present in at least one isolate (shared) on the same locus while the red blocks represent unique recombination events (different sizes and locations) that were not shared at a particular locus. (iii) Scale showing the position in the genome.



pneumococcal virulence [36]. Griffiths demonstrated that genetic recombination through a transforming factor, now known to be DNA [74] from heat inactivated virulent (encapsulated) pneumococci induced virulence in mice that were infected with avirulent (unencapsulated) pneumococcal strains. In addition to demonstrating the biological importance of recombination in virulence and disease, this discovery marked the beginning of the new era of molecular genetics [36].

Genetic recombination plays an important role in the development of antibiotic resistance in the pneumococcus. Antibiotic induced stress is known to induce competence in the pneumococcus; during the competence phase, the pneumococcus acquires exogenous DNA, which may include genes that confer resistance to antibiotics [75]. It has been reported that recombination replacements are responsible for the mosaic structure typically observed in penicillin binding proteins (PBP) genes in *S. pneumoniae* [76]. Mutations in PBP genes confer resistance to  $\beta$ -lactam antibiotics including penicillin, amoxicillin and cefotaxime [77].  $\beta$ -lactams kill bacteria by inhibiting cell wall biosynthesis and are still by far the most widely used class of antibiotics, as such the increasing rates of resistance in pneumococci are a major global health concern. Recombination with other oral streptococci, particularly the mitis and viridans group played a major role in the initial acquisition of  $\beta$ -lactam resistance in the pneumococcus [79,80]. Recombination involving PBP genes has also been extensively documented to occur between pneumococcal serotypes such as serotypes 9V $\leftrightarrow$ 23F, 9N $\leftrightarrow$ 14, 7B $\leftrightarrow$ 9N, 35C $\leftrightarrow$ 17F and 12F $\leftrightarrow$ 7F [68,78]. Recombination in *S. pneumoniae* isolates has also been associated with increased levels of resistance to multiple other classes of antibiotics [30]. Recombination mediates the dissemination of transposon and integrative conjugative elements (ICEs) that carry an array of antibiotic resistance determinants, throughout the pneumococcal population [81,82]. Such mobile genetic elements include the Tn916-like mobile genetic elements (MGEs), Tn5251 (a Tn916-like element) [83], Tn5252 [84], Tn5253 (a composite of Tn5251 and Tn5252 transposons) [84] and many others. The Tn916, Tn5251, Tn5252, Tn5253 and Tn1545 transposons carry *tetM* gene, which confers resistance to tetracycline [83–86]. In addition to the *tetM* gene, Tn5251 transposons (the Omega element) also carry the *catQ* gene that confers resistance to chloramphenicol. Another Tn916 family transposon, Tn1545, facilitates resistance to macrolide antibiotics (e.g. erythromycin and azithromycin), mediated by the ribosomal protection gene (*ermB*) and through efflux mechanism (*mefE*) gene that it carries [87]. Several epidemiological studies have reported associations between levels of antibiotic resistance in pneumococcal isolates and the presence of mobile genetic elements, which suggests that genetic recombination involving such MGEs has significant clinical impact [88].

Specific loci accumulate recombination events at a higher rate compared to others. Such regions are known as hotspots of recombination. A recent study of the largest collection of sequenced *S. pneumoniae* genomes to date ( $n = 3085$ ) found an association between genetic recombination hotspots and antibiotic resistance [51]. *S. pneumoniae* isolates that had undergone a recombination in the genes targeted by cotrimoxazole (*folA*) and  $\beta$ -lactam antibiotics (*pbp1a*, *pbp2a* and *pbp2x*) had higher likelihood to exhibit antibiotic resistance to the aforementioned antibiotics [51,89]. In contrast to *tetM* and *catQ* genes, acquired accessory genes, which confer resistance to tetracycline and chloramphenicol respectively, the *folA* and PBP genes are core housekeeping that are not disseminated by MGEs. Recombination encompassing the core genome encoded topoisomerase type II genes has been implicated in fluoroquinolones in both viridans streptococci [90] and salmonella [91], however its contribution to resistance in pneumococci has been reported to be minimal [92,93]. In PMEN1, recombination events and point mutations in the *rpoB* gene were also associated with rifampicin resistance [24]. In addition to *folA* and PBPs, other genes reported to be pneumococcal recombination hotspots, are the virulence factors and protein vaccine candidates, *pspA* and *pspC* (*cbpA*) [51]. *pspA* is a cell wall surface protein and a candidate for protein based vaccines and

has been shown to increase virulence in mice [94]. *pspC* (*cbpA* or *spcA*) is a choline-binding cell surface protein that plays a role in establishment of colonisation [95]. Thus, the occurrence of genetic recombination in these vaccine candidate and antibiotic resistance genes could result in successful evasion of both the humoral and cell mediated hosts' immune responses and increased levels of resistance to the targeted antibiotics.

## 7. Recombination Drives Vaccine Escape

Since the introduction of PCV7 in developed countries, followed by PCV10 and PCV13, a significant reduction in invasive pneumococcal disease has been reported [96]. PCVs directly work against a set of pneumococcal serotypes whose capsular polysaccharides have been targeted in the vaccine formulation. The introduction of PCVs has led to the emergence of the capsule-switch variants arising due to the vaccine-derived selective pressures [97]. Capsule switching is a natural process that occurs when different pneumococcal serotypes 'swap' their capsular polysaccharide locus through alteration of the capsule biosynthesis locus via mutations (single base changes, insertions or deletions) or genetic recombination. When a vaccine-type (VT) strain 'swaps' its capsule with a non-vaccine type (NVT) strain, it is able to escape the vaccine since PCVs only confer protection against the VT. An increase in the prevalence of these capsule-switched variants can result in serotype replacement, which is the increase of NVT associated pneumococcal clones, that follows the decrease in VT associated clones [68]. Such replacement of VTs by NVTs has the potential to reduce the impact of vaccination on the overall IPD burden in the long term. Various reports have documented the emergence and circulation of vaccine escape recombinants between serotype VT serotype 19F and NVT serotype 19A following the introduction of PCV7 in the United States [97]. Similar capsule switch variants were also reported in study cohorts from Massachusetts and Thailand [24,32,51]. Apart from capsule-switch, serotype replacement that occurs after introduction of PCVs is mainly caused by 'serotype unmasking' [98]. Unmasking is the process where less prevalent or 'masked' NVT serotypes rise in prevalence to occupy the ecological niche vacated by the 'more' competitive VT serotypes after vaccination [98]. Several reports have shown that serotype replacement is predominantly caused by the unmasking of less prevalent NVT serotypes. Analysis of the genetic background of the replacement serotypes is used to rule out whether such increase of NVT serotypes post-vaccination was caused by capsule switching. For example, comparing antibiotic susceptibility patterns and MLST sequence types (ST) of the replacement serotypes could determine whether or not the same genotypes were present in before the vaccine rollout [99]. Presence of the STs in the replacement serotypes that are associated with different serotypes could suggest that serotype switching has taken place through either point mutations or recombination [99].

## 8. Conclusions

The impact of genetic recombination to the adaptation and evolution of *S. pneumoniae* has been demonstrated by several studies. This has been largely due to parallel advancements in both analysis methods and high throughput sequencing technology. Such advances allow the analyses of whole genome sequences from diverse collections of pneumococcal isolates. Despite these successes, there are still several questions regarding the nature and biological impacts of recombination in pneumococci and other bacterial species. Further studies are required to uncover the exact functional roles of certain recombination events. Most of the genomic studies on pneumococcal recombination have been from a macroevolution standpoint whereby recombination replacements have been analysed at a population level and over large time scales. To better understand the biological roles of these recombination events, future genomic studies should aim to identify the

functional roles of certain recombination events in order to provide further insights into pneumococcal pathogenesis, nasopharyngeal carriage dynamics and strain transmission. This knowledge would be invaluable to the development of preventative and control strategies against this important pathogen.

## Acknowledgements

CC is a Commonwealth PhD Student at the Institute of Infection and Global Health (IGH), University of Liverpool funded by the United Kingdom government through the Commonwealth Scholarship Commission (CSC). The authors would like to thank Dr Nicholas J. Croucher for giving us permission to use a subset of the PMEN1 pneumococcal isolates for the figure included in the manuscript. We would also like to acknowledge support from the Malawi-Liverpool-Wellcome Trust Clinical Research Programme and the IGH, University of Liverpool.

## References

- [1] Osler W. The principles and practice of medicine. 1st ed. New York, NY: D. Appleton and Company; 1892.
- [2] Bogaert D, De Groot R, Hermans PW. *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *Lancet Infect Dis* 2004;4(3):144–54.
- [3] Regev-Yochay G, et al. Nasopharyngeal carriage of *Streptococcus pneumoniae* by adults and children in community and family settings. *Clin Infect Dis* 2004;38(5): 632–9.
- [4] Hill PC, et al. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Gambian villagers. *Clin Infect Dis* 2006;43(6):673–9.
- [5] Schuchat A, et al. Bacterial meningitis in the United States in 1995. Active Surveillance Team. *N Engl J Med* 1997;337(14):970–6.
- [6] Bohte R, van Furth R, van den Broek PJ. Aetiology of community-acquired pneumonia: a prospective study among adults requiring admission to hospital. *Thorax* 1995; 50(5):543–7.
- [7] O'Brien KL, et al. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet* 2009;374(9693):893–902.
- [8] Ortqvist A, Hedlund J, Kalin M. *Streptococcus pneumoniae*: epidemiology, risk factors, and clinical features. *Semin Respir Crit Care Med* 2005;26(6):563–74.
- [9] Bentley SD, et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet* 2006;2(3):e31.
- [10] Calix JJ, Nahm MH. A new pneumococcal serotype, 11E, has a variably inactivated wcjE gene. *J Infect Dis* 2010;202(1):29–38.
- [11] Jin P, et al. First report of putative *Streptococcus pneumoniae* serotype 6D among nasopharyngeal isolates from Fijian children. *J Infect Dis* 2009;200(9):1375–80.
- [12] Park IH, et al. Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. *J Clin Microbiol* 2007;45(4):1225–33.
- [13] Malley R, et al. Intranasal immunization with killed unencapsulated whole cells prevents colonization and invasive disease by capsulated pneumococci. *Infect Immun* 2001;69(8):4870–3.
- [14] Brueggemann AB, et al. Clonal relationships between invasive and carriage *Streptococcus pneumoniae* and serotype- and clone-specific differences in invasive disease potential. *J Infect Dis* 2003;187(9):1424–32.
- [15] Johnson HL, et al. Systematic evaluation of serotypes causing invasive pneumococcal disease among children under five: the pneumococcal global serotype project. *PLoS Med* 2010;7(10):e1000348.
- [16] Hsu HE, et al. Effect of pneumococcal conjugate vaccine on pneumococcal meningitis. *N Engl J Med* 2009;360(3):244–56.
- [17] Moore MR, et al. Effect of use of 13-valent pneumococcal conjugate vaccine in children on invasive pneumococcal disease in children and adults in the USA: analysis of multisite, population-based surveillance. *Lancet Infect Dis* 2015;15(3):301–9.
- [18] Hanage WP, et al. Clonal replacement among 19A *Streptococcus pneumoniae* in Massachusetts, prior to 13 valent conjugate vaccination. *Vaccine* 2011;29(48):8877–81.
- [19] Flasche S, et al. Effect of pneumococcal conjugate vaccination on serotype-specific carriage and invasive disease in England: a cross-sectional study. *PLoS Med* 2011; 8(4):e1001017.
- [20] Pilishvili T, et al. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. *J Infect Dis* 2010;201(1):32–41.
- [21] von Gottberg A, et al. Effects of vaccination on invasive pneumococcal disease in South Africa. *N Engl J Med* 2014;371(20):1889–99.
- [22] Hicks LA, et al. Incidence of pneumococcal disease due to non-pneumococcal conjugate vaccine (PCV7) serotypes in the United States during the era of widespread PCV7 vaccination, 1998–2004. *J Infect Dis* 2007;196(9):1346–54.
- [23] Dagan R, et al. Efficacy of 13-valent pneumococcal conjugate vaccine (PCV13) versus that of 7-valent PCV (PCV7) against nasopharyngeal colonization of antibiotic-nonsusceptible *Streptococcus pneumoniae*. *J Infect Dis* 2015;211(7):1144–53.
- [24] Croucher NJ, et al. Rapid pneumococcal evolution in response to clinical interventions. *Science* 2011;331(6016):430–4.
- [25] Brown T. Genomes. 2nd ed. Oxford: Wiley-Liss; 2002.
- [26] Darwin C. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. London: John Murray; 1859.
- [27] Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. *Genetics* 2000; 155(3):1405–13.
- [28] Bryant J, Chewapreecha C, Bentley SD. Developing insights into the mechanisms of evolution of bacterial pathogens from whole-genome sequences. *Future Microbiol* 2012;7(11):1283–96.
- [29] Vos M. Why do bacteria engage in homologous recombination? *Trends Microbiol* 2009;17(6):226–32.
- [30] Hanage WP, et al. Hyper-recombination, diversity, and antibiotic resistance in pneumococcus. *Science* 2009;324(5933):1454–7.
- [31] Castillo-Ramirez S, et al. Phylogeographic variation in recombination rates within a global clone of methicillin-resistant *Staphylococcus aureus*. *Genome Biol* 2012; 13(12):R126.
- [32] Croucher NJ, et al. Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat Genet* 2013;45(6):656–63.
- [33] Feil E, et al. The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. *Mol Biol Evol* 1999;16:1496–502.
- [34] Feil EJ, Enright MC, Spratt BG. Estimating the relative contributions of mutation and recombination to clonal diversification: a comparison between *Neisseria meningitidis* and *Streptococcus pneumoniae*. *Res Microbiol* 2000;151(6):465–9.
- [35] Shapiro BJ. Signatures of natural selection and ecological differentiation in microbial genomes. *Adv Exp Med Biol* 2014;781:339–59.
- [36] Griffith F. The significance of pneumococcal types. *J Hyg (Lond)* 1928;27(2):113–59.
- [37] Croucher NJ, et al. A high-resolution view of genome-wide pneumococcal transformation. *PLoS Pathog* 2012;8(6):e1002745.
- [38] Havarstein LS, Coomaraswamy G, Morrison DA. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. *Proc Natl Acad Sci U S A* 1995;92(24):11140–4.
- [39] Miller MB, Bassler BL. Quorum sensing in bacteria. *Annu Rev Microbiol* 2001;55: 165–99.
- [40] Claverys J-P, Prudhomme M, Martin B. Induction of competence regulons as a general response to stress in gram-positive bacteria. *Annu Rev Microbiol* 2006;60(1): 451–75.
- [41] Tomasz A. Control of the competent state in pneumococcus by a hormone-like cell product: an example for a new type of regulatory mechanism in bacteria. *Nature* 1965;208(5006):155–9.
- [42] Martin B, et al. Cross-regulation of competence pheromone production and export in the early control of transformation in *Streptococcus pneumoniae*. *Mol Microbiol* 2000;38(4):867–78.
- [43] Pestova EV, Havarstein LS, Morrison DA. Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. *Mol Microbiol* 1996;21(4): 853–62.
- [44] Peterson SN, et al. Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. *Mol Microbiol* 2004;51(4): 1051–70.
- [45] Claverys JP, Martin B, Havarstein LS. Competence-induced fratricide in streptococci. *Mol Microbiol* 2007;64(6):1423–33.
- [46] Smith MD, et al. Transfer of plasmids by conjugation in *Streptococcus pneumoniae*. *Plasmid* 1980;3(1):70–9.
- [47] Steinmoen H, Knutsen E, Havarstein LS. Induction of natural competence in *Streptococcus pneumoniae* triggers lysis and DNA release from a subfraction of the cell population. *Proc Natl Acad Sci* 2002;99(11):7681–6.
- [48] Hiller NL, et al. Generation of genic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal pediatric infection. *PLoS Pathog* 2010;6(9):e1001108.
- [49] Marks LR, Reddinger RM, Hakansson AP. High levels of genetic recombination during nasopharyngeal carriage and biofilm formation in *Streptococcus pneumoniae*. *MBio* 2012;3(5).
- [50] Magee AD, Yother J. Requirement for capsule in colonization by *Streptococcus pneumoniae*. *Infect Immun* 2001;69(6):3755–61.
- [51] Chewapreecha C, et al. Dense genomic sampling identifies highways of pneumococcal recombination. *Nat Genet* 2014;46(3):305–9.
- [52] Croucher NJ, et al. Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone *Streptococcus pneumoniae* Spain23F ST81. *J Bacteriol* 2009;191(5):1480–9.
- [53] King SJ, Whatmore AM, Dowson CG. NanaA, a neuraminidase from *Streptococcus pneumoniae*, shows high levels of sequence diversity, at least in part through recombination with *Streptococcus oralis*. *J Bacteriol* 2005;187(15):5376–86.
- [54] Sung CK, et al. An rpsL cassette, janus, for gene replacement through negative selection in *Streptococcus pneumoniae*. *Appl Environ Microbiol* 2001;67(11): 5190–6.
- [55] Donati C, et al. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. *Genome Biol* 2010;11(10):R107.
- [56] Tettelin H, et al. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* 2001;293:498–506.
- [57] Bentley S. Sequencing the species pan-genome. *Nat Rev Microbiol* 2009;7(4):258–9.
- [58] Jackson RW, et al. The influence of the accessory genome on bacterial pathogen evolution. *Mob Genet Elem* 2011;1(1):55–65.
- [59] Posada D, Crandall KA, Holmes EC. Recombination in evolutionary genomics. *Annu Rev Genet* 2002;36:75–97.
- [60] Didelot X, et al. Inference of homologous recombination in bacteria using whole-genome sequences. *Genetics* 2010;186(4):1435–49.
- [61] Marttinen P, et al. Detection of recombination events in bacterial genomes from large population samples. *Nucleic Acids Res* 2012;40(1):e6.
- [62] Hakenbeck R, et al. Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of *Streptococcus pneumoniae*. *Infect Immun* 2001;69:2477–86.

- [63] Croucher NJ, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 2015;43(3):e15.
- [64] Harris SR, et al. Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. *Nat Genet* 2012;44(4):413–9 [S1].
- [65] Fraser C, Hanage WP, Spratt BG. Recombination and the nature of bacterial speciation. *Science* 2007;315(5811):476–80.
- [66] Cooper TF. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. *PLoS Biol* 2007;5(9):e225.
- [67] Lang GI, et al. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* 2013;500(7464):571–4.
- [68] Wyres KL, et al. Pneumococcal capsular switching: a historical perspective. *J Infect Dis* 2013;207(3):439–49.
- [69] Andam CP, Hanage WP. Mechanisms of genome evolution of *Streptococcus*. *Infection, Genetics and Evolution* 2014 Nov 13, pii: S1567-1348(14)00410-9.
- [70] Coffey TJ, et al. Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of *Streptococcus pneumoniae*. *Mol Microbiol* 1998;27(1):73–83.
- [71] Croucher NJ, et al. Dominant role of nucleotide substitution in the diversification of serotype 3 pneumococci over decades and during a single infection. *PLoS Genet* 2013;9(10):e1003868.
- [72] Mostowy R, et al. Heterogeneity in the frequency and characteristics of homologous recombination in pneumococcal evolution. *PLoS Genet* 2014;10(5):e1004300.
- [73] McGee L, et al. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the pneumococcal molecular epidemiology network. *J Clin Microbiol* 2001;39(7):2565–71.
- [74] Avery OT, Macleod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J Exp Med* 1944;79(2):137–58.
- [75] Slager J, et al. Antibiotic-induced replication stress triggers bacterial competence by increasing gene dosage near the origin. *Cell* 2014;157(2):395–406.
- [76] Dowson C, et al. Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. *Proc Natl Acad Sci U S A* 1989;86:8842–6.
- [77] Hakenbeck R, et al. Beta-lactam resistance in *Streptococcus pneumoniae*: penicillin-binding proteins and non-penicillin-binding proteins. *Mol Microbiol* 1999;33:673–8.
- [78] Coffey TJ, et al. Serotype 14 variants of the Spanish penicillin-resistant serotype 9 V clone of *Streptococcus pneumoniae* arose by large recombinational replacements of the *cpsA-pbp1a* region. *Microbiology* 1999;145(Pt 8):2023–31.
- [79] Dowson C, et al. Penicillin-resistant viridans streptococci have obtained altered penicillin-binding protein genes from penicillin-resistant strains of *Streptococcus pneumoniae*. *Proc Natl Acad Sci U S A* 1990;87:5858–62.
- [80] Dowson CG, et al. Evolution of penicillin resistance in *Streptococcus pneumoniae*; the role of *Streptococcus mitis* in the formation of a low affinity PBP2B in *S. pneumoniae*. *Mol Microbiol* 1993;9(3):635–43.
- [81] Caparon MG, Scott JR. Excision and insertion of the conjugative transposon Tn916 involves a novel recombination mechanism. *Cell* 1989;59(6):1027–34.
- [82] Cornick JE, Bentley SD. *Streptococcus pneumoniae*: the evolution of antimicrobial resistance to beta-lactams, fluoroquinolones and macrolides. *Microbes Infect* 2012;14(7–8):573–83.
- [83] Provvedi R, Manganello R, Pozzi G. Characterization of conjugative transposon Tn5251 of *Streptococcus pneumoniae*. *FEMS Microbiol Lett* 1996;135(2–3):231–6.
- [84] Ayoubi P, Kilic AO, Vijayakumar MN. Tn5253, the pneumococcal omega (cat tet) BM6001 element, is a composite structure of two conjugative transposons, Tn5251 and Tn5252. *J Bacteriol* 1991;173(5):1617–22.
- [85] Martin P, Trieu-Cuot P, Courvalin P. Nucleotide sequence of the tetM tetracycline resistance determinant of the streptococcal conjugative shuttle transposon Tn1545. *Nucleic Acids Res* 1986;14(17):7047–58.
- [86] Rice LB. Tn916 family conjugative transposons and dissemination of antimicrobial resistance determinants. *Antimicrob Agents Chemother* 1998;42(8):1871–7.
- [87] Okitsu N, et al. Characterization of ermB gene transposition by Tn1545 and Tn917 in macrolide-resistant *Streptococcus pneumoniae* isolates. *J Clin Microbiol* 2005;43(1):168–73.
- [88] Shiojima T, et al. Prevalence of *Streptococcus pneumoniae* isolates bearing macrolide resistance genes in association with integrase genes of conjugative transposons in Japan. *Clin Microbiol Infect* 2005;11(10):808–13.
- [89] Cornick JE, et al. Genomic identification of a novel co-trimoxazole resistance genotype and its prevalence amongst *Streptococcus pneumoniae* in Malawi. *J Antimicrob Chemother* 2013;69(2):368–74.
- [90] Pletz MW, et al. Fluoroquinolone resistance in invasive *Streptococcus pyogenes* isolates due to spontaneous mutation and horizontal gene transfer. *Antimicrob Agents Chemother* 2006;50(3):943–8.
- [91] Turner AK, Nair S, Wain J. The acquisition of full fluoroquinolone resistance in *Salmonella Typhi* by accumulation of point mutations in the topoisomerase targets. *J Antimicrob Chemother* 2006;58(4):733–40.
- [92] Bast DJ, et al. Interspecies recombination contributes minimally to fluoroquinolone resistance in *Streptococcus pneumoniae*. *Antimicrob Agents Chemother* 2001;45(9):2631–4.
- [93] Balsalobre L, et al. Viridans group streptococci are donors in horizontal transfer of topoisomerase IV genes to *Streptococcus pneumoniae*. *Antimicrob Agents Chemother* 2003;47(7):2072–81.
- [94] McCool TL, et al. The immune response to pneumococcal proteins during experimental human carriage. *J Exp Med* 2002;195(3):359–65.
- [95] Rosenow C, et al. Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of *Streptococcus pneumoniae*. *Mol Microbiol* 1997;25(5):819–29.
- [96] Whitney C, et al. Decline in invasive pneumococcal disease after the introduction of protein-polysaccharide conjugate vaccine. *N Engl J Med* 2003;348:1737–46.
- [97] Brueggemann AB, et al. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS Pathog* 2007;3(11):e168.
- [98] Lipsitch M. Bacterial vaccines and serotype replacement: lessons from *Haemophilus influenzae* and prospects for *Streptococcus pneumoniae*. *Emerg Infect Dis* 1999;5(3):336–45.
- [99] Pai R, et al. Postvaccine genetic structure of *Streptococcus pneumoniae* serotype 19A from children in the United States. *J Infect Dis* 2005;192(11):1988–95.